(19) JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: 10260979 A

(43) Date of publication of application: 29.09.98

(51) Int. Cl G06F 17/30

(21) Application number: 09065106

(22) Date of filing: 18.03.97

(71) Applicant: NIPPON TELEGR & TELEPH CORP <NTT>

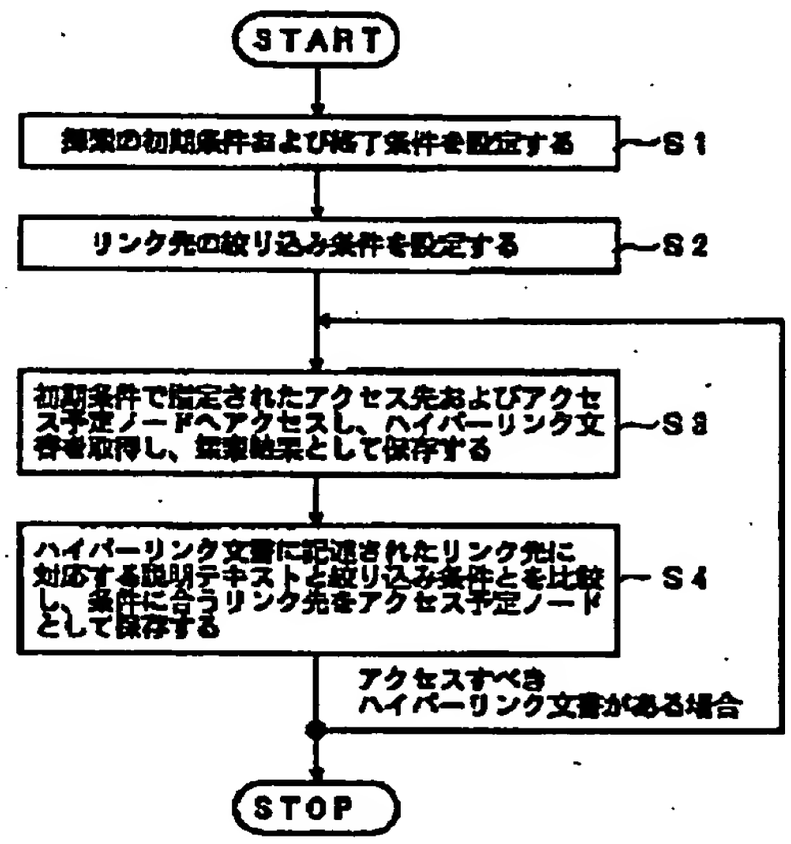
(72) Inventor: MATSUO HIROSHI

(54) INFORMATION COLLECTING METHOD AND DEVICE

(57) Abstract:

PROBLEM TO BE SOLVED: To efficiently collect information by comparing an explanation text that corresponds to a link destination described on a hyperlink document with a narrowing condition and holding the link destination that coincides with the narrowing condition as an access prearranged node.

SOLUTION: A start condition of search and an end condition of search are set (step S1), a narrowing condition of a link destination is set (step S2), a hyperlink document is acquired by accessing an access destination that is designated with the start condition and an access prearranged node, and the acquired hyperlink document is stored as a search result (step S3). An explanation text that corresponds to a link destination which is described on the hyperlink document is compared with a narrowing condition, and a link destination that coincides with the narrowing condition is held as an access prearranged node (step S4). The above processing is performed of entire hyperlink document to be accessed.



1

【特許請求の範囲】

【請求項１】　ネットワーク上に分散されたサーバ上にハイパーリンク文書が分散蓄積されている場合の情報収集を支援するための情報収集方法において、
探索の初期条件及び終了条件を設定し、
リンク先の絞り込み条件を設定し、
前記初期条件で指定されたアクセス先及びアクセス予定ノードへアクセスし、前記ハイパーリンク文書を取得し、当該ハイパーリンク文書を探索結果として保持しておき、
前記ハイパーリンク文書に記述されたリンク先に対応する説明テキストと前記絞り込み条件とを比較し、該絞り込み条件に合致するリンク先をアクセス予定ノードとして保持する処理を、アクセスすべき全ハイパーリンク文書について行うことを特徴とする情報収集方法。

【請求項２】　前記リンク先の絞り込み条件として指定されたキーワードを、語の概念の関係を表すシソーラス辞書に基づいて展開し、照合対象語として設定し、
前記ハイパーリンク文書に記述されたリンク先に対応する説明テキスト中に、前記照合対象語が含まれるリンク先をアクセス予定ノードとする請求項１記載の情報収集方法。

【請求項３】　ハイパーリンク文書のリンク先を辿って文書の自動収集を行う情報収集装置であって、
探索の初期条件及び終了条件を設定する探索条件設定手段と、
リンク先の絞り込み条件を設定する絞り込み条件設定手段と、
ハイパーリンク文書に記述されたリンク先に対応する説明テキストを前記絞り込み条件に基づいて判定し、条件に合うリンク先をアクセス予定ノードとするアクセス候補絞り込み手段と、
前記初期条件で指定されたアクセス先及び前記アクセス予定ノードへアクセスしてハイパーリンク文書を取得し、該ハイパーリンク文書を探索結果として保存するハイパーリンク文書アクセス手段と、
前記初期条件及び前記終了条件に基づいて、前記ハイパーリンク文書アクセス手段及び前記アクセス候補絞り込み手段を繰り返し起動する探索管理手段とを有することを特徴とする情報収集装置。

【請求項４】　語の概念の関係を表すシソーラス辞書を更に有し、
前記アクセス候補絞り込み手段は、
前記絞り込み条件として指定されたキーワードを、前記シソーラス辞書に基づいて展開して照合対象語として設定する絞り込み条件設定手段と、
前記ハイパーリンク文書の記述されたリンク先に対応する説明テキスト中に前記照合対象語が含まれるリンク先をアクセス予定ノードとする絞り込み手段を含む請求項３記載の情報収集装置。

2

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、情報収集方法及び装置に係り、特に、ハイパーリンク文書のリンク先を辿って文書の自動収集を行う情報収集方法及び装置に関する。詳しくは、ネットワーク上に分散されたサーバ上にハイパーリンク文書が分散蓄積されている場合の情報収集を支援するための情報収集方法及び装置に関する。

【０００２】

【従来の技術】従来の情報収集装置について説明する。図８は、従来の情報収集装置の構成を示す。従来の情報収集装置は、探索条件設定部Ｃ１、探索部Ｃ２及び探索結果表示部Ｃ３から構成される。探索部Ｃ２は、ハイパーリンク文書アクセス部Ｃ２１，アクセス予定ノード抽出部Ｃ２２、終了条件判定部Ｃ２３より構成される。

【０００３】図９は、従来の情報収集装置の動作を示す。ステップ１０）　まず、探索条件設定部Ｃ１により、初期アクセスノードと最大ノード深さを探索条件として設定する。ステップ１１）　初期条件として、初期アクセスノードをカレントノードとする。

【０００４】ステップ１２）　ハイパーリンク文書アクセス部Ｃ２１により、カレントノードにアクセスし、ハイパーリンク文書を取得し、探索結果として保存する。ステップ１３）　終了条件判定部Ｃ２３により、カレントノードが最大深さにあるかを調べ、最大深さにある場合には、次のステップ１５に移行する。最大深さにない場合には、ステップ１４に移行する。

【０００５】ステップ１４）　アクセス予定ノード抽出部Ｃ２２により、ハイパーリンク文書に記述されたリンク先をアクセス予定ノードとして保存する。ステップ１５）　終了条件判定部Ｃ２３により、未処理のアクセス予定ノードがあるかを調べ、あれば、それをカレントノードとする。未処理のアクセス予定ノードがある場合には、ステップ１２に移行する。

【０００６】ステップ１６）　探索結果表示部Ｃ３により探索結果を表示する。以上の処理により、初期アクセスノードから最大ノード深さまでのリンクされたハイパーリンク文書を全て取得することができる。

【０００７】

【発明が解決しようとする課題】しかしながら、上記従来の装置では、リンクされたすべてのハイパーリンク文書にアクセスするため、指定する最大ノード深さが深い場合には、大量の文書を取得してしまう。１文書あたりｎ個のリンク先があり、最大ノード深さをｍと指定した場合、ｎのｍ乗個となり、探索の深さを大きくすると、膨大な量の文書をアクセスすることになる。例えば、ｎ＝５，ｍ＝５とすると、５階層目で３１２５個の文書をアクセスすることになる。

【０００８】しかしながら、ユーザが欲しい情報は、これら文書のうちごく一部であることが多い。このため、

検索処理を施して収集した文書を絞り込んだり、人手で必要な文書を捜し出すなどの作業に多くの労力を要する。また、ネットワーク上に分散配置された文書を大量にアクセスすると、ネットワークへの大きな負担がかかるという問題がある。

【０００９】本発明は、上記の点に鑑みなされたもので、すべてのリンク先をアクセスするのではなく、絞り込み条件を満たすリンク先のみをアクセスすることによって、ユーザの所望する可能性が高い文書に対するアクセスに限定して、効率よく情報を収集することが可能な情報収集方法及び装置を提供することを目的とする。

【００１０】

【課題を解決するための手段】図１は、本発明の原理を説明するための図である。本発明は、ネットワーク上に分散されたサーバ上にハイパーリンク文書が分散蓄積されている場合の情報収集を支援するための情報収集方法において、探索の初期条件及び終了条件を設定し（ステップ１）、リンク先の絞り込み条件を設定し（ステップ２）、初期条件で指定されたアクセス先及びアクセス予定ノードへアクセスし、ハイパーリンク文書を取得し、取得したハイパーリンク文書を探索結果としての保存を行い（ステップ３）、ハイパーリンク文書に記述されたリンク先に対応する説明テキストと絞り込み条件とを比較し、該絞り込み条件に合致するリンク先をアクセス予定ノードとして保持する（ステップ４）処理を、アクセスすべき全ハイパーリンク文書について行う。

【００１１】また、本発明は、リンク先の絞り込み条件として指定されたキーワードを、語の概念の関係を表すシソーラス辞書に基づいて展開し、照合対象語として設定し、ハイパーリンク文書に記述されたリンク先に対応する説明テキスト中に、照合対象語が含まれるリンク先をアクセス予定ノードとする。

【００１２】図２は、本発明の原理構成図である。本発明は、ハイパーリンク文書のリンク先を辿って文書の自動収集を行う情報収集装置であって、探索の初期条件及び終了条件を設定する探索条件設定手段１と、リンク先の絞り込み条件を設定する絞り込み条件設定手段２と、ハイパーリンク文書に記述されたリンク先に対応する説明テキストを絞り込み条件に基づいて判定し、絞り込み条件に合うリンク先をアクセス予定ノードとするアクセス候補絞り込み手段３と、初期条件で指定されたアクセス先及びアクセス予定ノードへアクセスしてハイパーリンク文書を取得し、該ハイパーリンク文書を探索結果として保存するハイパーリンク文書アクセス手段４と、初期条件及び終了条件に基づいて、ハイパーリンク文書アクセス手段４及びアクセス候補絞り込み手段３を繰り返し起動する探索管理手段５とを有する。

【００１３】また、本発明において、語の概念の関係を表すシソーラス辞書を更に有し、アクセス候補絞り込み手段３は、絞り込み条件として指定されたキーワードを、シソーラス辞書に基づいて展開して照合対象語として設定する絞り込み条件設定手段と、ハイパーリンク文書の記述されたリンク先に対応する説明テキスト中に照合対象語が含まれるリンク先をアクセス予定ノードとする絞り込み手段を含む。

【００１４】このように、本発明による情報収集方法及び装置では、アクセスしたハイパーリンク文書に記述されたリンク先に対応する説明テキストと、絞り込み条件設定手段により設定されている絞り込み条件とを比較することにより、合致するリンク先をアクセス予定ノードとして絞り込むことにより、不要な文書へのアクセスを回避して、ユーザが所望する情報が含まれる可能性がある文書を効率よく収集することが可能となる。

【００１５】

【発明の実施の形態】図３は、本発明の情報収集装置の構成を示す。同図における情報収集装置は、探索条件設定部１、絞り込み条件設定部２、アクセス候補絞り込み部３、ハイパーリンク文書アクセス部４、探索管理部５、アクセス管理テーブル６、取得文書蓄積部７、処理要求受付部８、探索結果表示部９、照合対象語蓄積部１０、及びシソーラス辞書１１から構成され、当該情報収集装置は、ネットワーク１２を介してサーバ群１３に接続される。

【００１６】アクセス管理テーブル６は、アクセス予定の文書の識別子（以下、アクセス予定ノードと呼ぶ）等を保存して、文書のアクセスの管理に使用するテーブルである。当該アクセス管理テーブル６には、アクセス予定ノード、処理結果、ノードの深さ等の項目が定義される。アクセス予定ノードには、アクセス予定として決定された文書の識別子が記述され、処理結果には、当該文書が取得されたか否かによって、『処理済』か『未処理』が記述され、ノード深さには、開始文書を０として、開始文書から何個のリンクで結ばれているかを示す個数が記述される。

【００１７】取得文書蓄積部７は、ハイパーリンク文書アクセス部４によって取得された文書を蓄積しておく。処理要求受付部８は、ユーザとの入出力インタフェースであり、情報入出力の機能を持つか、ネットワークを介して端末と接続され、ユーザからの探索条件設定要求、絞り込み条件設定要求、探索要求、探索結果出力要求を受付、各々、探索条件設定部１、絞り込み条件設定部２、探索管理部５、探索結果表示部９へ仲介する。

【００１８】探索結果表示部９は、処理要求受付部８からの要求に応じて、取得文書蓄積部７に蓄積された文書を出力する。照合対象語蓄積部１０は、アクセス候補絞り込み部３で参照できるように、絞り込み条件設定部２によって設定される照合対象語を保存しておく。シソーラス辞書１１は、語が表す概念の関係を示す情報を蓄積した辞書であり、絞り込みを行うための知識として使用される。

【００１９】探索条件設定部１は、処理要求受付部８を介して探索条件の要求を受け取り、探索条件の初期条件として開始文書を設定し、終了条件として、最大ノード深さを設定する。絞り込み条件設定部２は、処理要求受付部８を介して絞り込み条件設定要求を受け取り、絞り込み条件を設定する。絞り込み条件の指定の方法として、キーワードを指定する方法、条件式を指定する方法、照合パターンを指定する方法等、各種指定方法を適用することができる。

【００２０】探索管理部５は、処理要求受付部８を介して探索要求を受け取ると、アクセス管理テーブル６を参照・更新しながら、探索条件で指定された終了条件に達するまで、アクセス候補絞り込み部３と、ハイパーリンク文書アクセス部４とを繰り返し起動して探索を行う。アクセス候補絞り込み部３は、ハイパーリンク文書中から、リンク先に対応する説明テキストを抽出し、照合対象語蓄積部１０に保存された照合対象語が説明テキスト中に含まれるリンク先をアクセス予定ノードとして抽出し、探索管理部５を介して、アクセス管理テーブル６へ登録する。

【００２１】ハイパーリンク文書アクセス部４は、探索管理部５からアクセスすべきノードの情報を受け取り、ネットワーク１２を介して、サーバ群１３をアクセスして、該当するハイパーリンク文書を取得し、取得文書蓄積部７へ保存する。図４は、本発明の情報収集方法の一連の動作を示すフローチャートである。

ステップ１０１）　探索条件設定部１は、探索条件の初期条件として、開始文書が設定され、終了条件として最大ノード深さが設定される。

【００２２】ステップ１０２）　アクセス候補絞り込み部３は、開始文書をアクセス予定ノードとして、アクセス管理テーブル６へ登録する。

ステップ１０３）　絞り込み条件設定部２は、処理要求受付部８からの絞り込み条件を受け取り、キーワードが絞り込み条件として設定される。

ステップ１０４）　絞り込み条件設定部２において、絞り込み条件で指定されたキーワードのシソーラス辞書１１上の位置を取得し、絞り込み条件設定部５においてその同義語、上位語、下位語を抽出し、照合対象語とし、照合対象語蓄積部１０に登録する。

【００２３】ステップ１０５）　ハイパーリンク文書アクセス部４は、探索管理部５を介して、アクセス管理テーブル６に登録されたアクセス予定ノードへアクセスし、ハイパーリンク文書を取得して、取得文書蓄積部７に探索結果として保存すると共に、探索管理部５を介して、アクセス管理テーブル６の当該アクセス予定ノードの＜処理結果＞を『処理済』とする。

【００２４】ステップ１０６）　探索管理部５は、取得した文書のノードの深さが最大ノード深さに達したかを調べ、最大ノード深さの場合には、ステップ１０９に移

行する。そうでない場合には、ステップ１０７に移行する。

ステップ１０７）　探索管理部５は、ハイパーリンク文書中から、リンク先に対応する説明テキストを抽出する。

【００２５】ステップ１０８）　アクセス候補絞り込み部３は、説明テキスト中に照合対象語が存在するかを調べ、存在する場合には、当該説明テキストに対応するリンク先をアクセス予定ノードとしてアクセス管理テーブル６へ登録する。

ステップ１０９）　探索管理部５は、アクセス管理テーブル中に『未処理』のアクセス予定ノードがあるか調べ、ある場合には、ステップ１０５に移行する。

【００２６】ステップ１１０）　ない場合には、探索結果表示部９において、探索結果出力要求を受け取ると、取得文書蓄積部７に蓄積された探索結果を出力する。

【００２７】

【実施例】以下、図面と共に本発明の実施例を説明する。図５は、本発明の一実施例のハイパーリンク文書の例を示す。図５に示す、ハイパーリンク文書の〔ｈ０００１〕，〔ｈ０００２〕等は、ハイパーリンク文書に付与した識別子である。ここで示した文書（以下、ここでは、ハイパーリンク文書を省略して単に文書と記す）は、説明のため簡略化しており、表や画像等を含んだ文書であってもよい。ハイパーリンク文書の場合、表示書式等の指定を行うタグやリンク先を示すためのタグが定義されており、これらの情報を基にリンク先に対応する説明テキストを抽出できる。また、自然言語解析処理を行って、タグ情報のみでは抽出できない説明テキストに対しても抽出を行う構成としてもよい。

【００２８】文書〔ｈ０００１〕では、文書〔ｈ００１１〕へのリンクに対し、「コンピュータ」が説明テキストに対応しており、文書〔ｈ００１２〕へのリンクに対し、「通販」が説明テキストに対応している。同様に、「企業」、「アート」、「メディア」は、各々〔ｈ００１３〕，〔ｈ００１４〕，〔ｈ００１５〕のリンクの説明テキストである。リンク先とそれに対応するテキストとを、

（＜リンク先＞、＜説明テキスト＞）

の形式で表すと、文書〔ｈ０００１〕には、リンク情報として、

（ｈ００１１，コンピュータ）
（ｈ００１２，通販）
（ｈ００１３，企業）
（ｈ００１４，アート）
（ｈ００１５，メディア）

が存在する。

【００２９】図６は、本発明の一実施例のシソーラス辞書の例を示す。同図に示すシソーラス辞書１１は、語が表す概念の関係を示す情報を蓄積した辞書であり、絞り

込みを行うための知識として使用される。同図の例では、両方向の矢印で同義語を表し、一方向の矢印で、上位・下位関係を表している。例えば、「販売」、「ショッピング」、「shopping」は、同義語であり、これらの語の下位語は、「通販」、「小売り」である。「小売り」から見ると、上位語は「販売」であり、下位語は、「百貨店」、「専門店」である。

【００３０】ここでは、キーワードを指定して絞り込み条件とする一例として、シソーラス辞書１１を参照して照合対象語に展開し、この照合対象語が説明テキスト中に含まれるか否かによって、リンク先を絞り込むよう構成した場合を例に説明する。本実施例では、アクセス候補絞り込み部３での参照が容易となるよう、絞り込み条件設定部２において、指定されたキーワードの同義語、上位語、下位語を照合対象語として照合対象蓄積部１０へ保持する。

【００３１】図７は、本発明の一実施例のアクセス管理テーブルの例を示す。同図（ａ）は、文書アクセス前のテーブルの状態を示し、（ｂ）は、〔ｈ０００１〕文書取得後のテーブルの状態を示し、（ｃ）は、〔ｈ００１２〕文書取得後のテーブルの状態を示す。以下、図４のフローチャートに従って、文書〔ｈ０００１〕が開始文書として設定され、最大ノード深さとして、“２”が設定された場合について説明する。

【００３２】ステップ１０１）　処理要求受付部８より、探索条件設定の要求を受け取り、探索条件の初期条件として、開始文書が設定され、終了条件として、最大ノード深さが設定される。即ち、開始文書として文書〔ｈ０００１〕が設定され、最大ノード深さとして“２”が設定される。

ステップ１０２）　探索の初期条件として指定された開始文書をアクセス予定ノードとしてアクセス管理テーブル６に登録する。指定される開始文書は、複数個であってもよい。ここでは、図５の文書〔ｈ０００１〕が開始文書として指定された場合について説明する。開始文書〔ｈ０００１〕に対しては、図７（ａ）に示すように、＜アクセス予定ノード＞の項に「ｈ０００１」が、＜処理結果＞の項に「未処理」が、＜ノードの深さ＞の項に「０」が記述される。

【００３３】ステップ１０３）　絞り込み条件設定部２において、処理要求受付部８からの、絞り込み条件設定要求を受け取り、キーワードが絞り込み条件として指定される。ここでは、キーワードとして、「ショッピング」と「家具」が指定されたものとする。

ステップ１０４）　絞り込み条件設定部２において、シソーラス辞書１１を参照し、絞り込み条件で指定されたキーワードの同義語、上位語、下位語を抽出し、照合対象語とし、これを照合対象語蓄積部１０へ保存する。図６のシソーラス辞書を使用した場合には、以下のように照合対象語が決定される。

【００３４】「ショッピング」に対する照合対象語としては、同義語として、「販売」、「ショッピング」、「shopping」が抽出され、下位語として「通販」、「小売り」が抽出され、「小売り」の下位語として「百貨店」、「専門店」が抽出される。「家具」に対しては、上位語として、「商品」が、下位語として、「机」、「椅子」が抽出される。

【００３５】即ち、照合対象語は、「販売」、「ショッピング」、「shopping」、「通販」、「小売り」、「百貨店」、「専門店」、「商品」、「机」、「椅子」となり、これらの語が照合対象語蓄積部１０に蓄積される。

ステップ１０５）　アクセス管理テーブル６に登録されたアクセス予定ノードにおいて、＜処理結果＞が『未処理』のノードへネットワーク１２を介してアクセスし、ハイパーリンク文書を取得する。そして、探索結果として取得文書蓄積部７へ取得したハイパーリンク文書を保存すると共に、当該アクセス予定ノードの＜処理結果＞を『処理済』とする。ここでは、説明の簡略化のため、リンク先をハイパーリンク文書の識別子のみで表しているが、リンク先としてハイパーリンク文書の所在を表す情報、即ち、どのサーバに存在するかという情報も含んでおり、どのサーバへアクセスすればよいかは決定できる。

【００３６】ステップ１０６）　取得した文書のノード深さが最大ノード深さに達したかを調べる。アクセス管理テーブル６の＜ノード深さ＞を検査することにより、最大ノード深さに達したかどうかを調べることができる。最大ノード深さに達した場合には、ステップ１０７、ステップ１０８をスキップして、ステップ１０９に移行する。

【００３７】ステップ１０７）　取得したハイパーリンク文書中から、リンク先に対応する説明テキストを抽出する。文書〔ｈ０００１〕の場合には、（ｈ００１１，コンピュータ）、（ｈ００１２，通販）、（ｈ００１３，企業）、（ｈ００１４，アート）、（ｈ００１５，メディア）が抽出される。

【００３８】ステップ１０８）　説明テキスト中に照合対象語が存在するかを調べ、存在する場合には、当該説明テキストに対応するリンク先をアクセス予定ノードとしてアクセス管理テーブル６へ登録する。文書〔ｈ００１〕の場合には、「通販」のみが、（ｈ００１２，通販）として存在する。そこで、文書〔ｈ００１２〕をアクセス管理テーブル６へ登録する。このとき、＜処理結果＞に『未処理』を記述し、＜ノードの深さ＞は親のノードの深さに１を加算したものを記述する。文書〔ｈ０００１〕の＜ノードの深さ＞は、０であるので、文書〔ｈ００１２〕の＜ノードの深さ＞は１となる。その結果、アクセス管理テーブル６の内容は、図７（ｂ）に示す通りである。

【００３９】ステップ１０９）　アクセス管理テーブル

６中に、『未処理』のアクセス予定ノードがあるかを調べ、ある場合にはステップ１０５に移行し、『未処理』のアクセス予定ノードがない場合には、ステップ１１０へ移行する。ステップ１１０）　処理要求受付部８より、探索結果出力要求を受け取ると、探索結果表示部９は、取得文書蓄積部７に蓄積された探索結果を処理要求受付部８に出力する。

【００４０】上記で説明した処理により、図５に示すハイパーリンク文書の場合には、ステップ１０５からステップ１１０に至るまでの処理は、以下のようになる。

（１）　初期条件の開始文書として指定された文書〔ｈ０００１〕が取得される（ステップ１０５）。

（２）　文書〔ｈ０００１〕の<ノード深さ>は、０であり、最大ノード深さ２に達していないため、次のステップ１０７に移行する（ステップ１０６）。

【００４１】（３）　文書〔ｈ０００１〕から、説明テキスト（ｈ００１１，コンピュータ）、（ｈ００１２，通販）、（ｈ００１３、企業）、（ｈ００１４，アート）、（ｈ００１５，メディア）が抽出される（ステップ１０７）。

【００４２】（４）　照合対象語の「通販」が（ｈ００１２，通販）に存在するので、〔ｈ００１２〕をアクセス予定ノードとしてアクセス管理テーブル６に登録する。

（５）　〔ｈ００１２〕が未処理なので、ステップ１０５に移行する（ステップ１０９）。

（６）　文書〔ｈ００１２〕が取得される（ステップ１０５）。

【００４３】（７）　文書〔ｈ００１２〕の<ノード深さ>は１であり、最大ノード深さ２に達していないため、次のステップ１０７に移行する（ステップ１０６）。

（８）　文書〔ｈ００１２〕から、説明テキスト（ｈ０１２１，百貨店）、（ｈ０１２２，ショッピングモール）、（ｈ０１２３，食品）、（ｈ０１２４，コンピュータ）、（ｈ０１２５，書籍）が抽出される（ステップ１０７）。

【００４４】（９）　照合対象語「百貨店」が（ｈ０１２１，百貨店）に存在し、照合対象語「ショッピング」は（ｈ０１２２，ショッピングモール）の文字列“ショッピングモール”中に含まれているので、〔ｈ０１２１〕及び〔ｈ０１２２〕がアクセス予定ノードとしてアクセス管理テーブル６に登録される。ここで、〔ｈ０１２１〕及び〔ｈ０１２２〕が未処理なので、ステップ１０５へ移行する（ステップ１０９）。

【００４５】（１０）　〔ｈ０１２１〕及び〔ｈ０１２２〕が未処理なので、ステップ１０５に移行する（ステップ１０９）。

（１１）　文書〔ｈ０１２１〕が取得される（ステップ１０５）。

（１２）　文書〔ｈ０１２１〕の<ノード深さ>は“２”であり、最大ノード深さ“２”に達しているため、ステップ１０７、ステップ１０８をスキップして、ステップ１０９に移行する（ステップ１０６）。

【００４６】（１３）　〔ｈ０１２２〕が『未処理』として残っているので、ステップ１０５に移行する（ステップ１０９）。

（１４）　文書〔ｈ０１２２〕が取得される（ステップ１０５）。

（１５）　文書〔ｈ０１２２〕の<ノード深さ>は、“２”であり、最大ノード深さ“２”に達しているため、ステップ１０７、ステップ１０８をスキップして、ステップ１０９に処理が移行する（ステップ１０６）。

【００４７】（１６）　未処理として残っているアクセス予定ノードは存在しないので、ステップ１１０に移行する（ステップ１０９）。

（１７）　探索結果出力の要求があると、取得文書蓄積部７に蓄積された文書〔ｈ０００１〕，〔ｈ００１２〕，〔ｈ０１２１〕，〔ｈ０１２２〕が出力される。

【００４８】このようにして、絞り込まれたリンク先のみへアクセスが行われる。上記の例のように、例えば、「家具を買いたい」という動機によって、「ショッピング」及び「家具」を指定すれば、ユーザが所望する情報が含まれる可能性のある文書以外へのアクセスが抑制される。このため、最大ノード深さを大きくとっても合理的な時間で、目的にあった情報を効率的に収集できる。

【００４９】なお、本発明は、上記の実施例に限定されることなく、特許請求の範囲内で種々変更・応用が可能である。

【００５０】

【発明の効果】上述のように、本発明の情報収集方法及び装置によれば、絞り込み条件でリンク先を絞り込むため、アクセスする文書量を大幅に削減でき、所望する情報が含まれる可能性が高い文書を容易に取得することができる。このため、本発明による装置を用いることにより、効率的に情報収集を行うことができる。

【００５１】従来の装置との比較を定量的に行うと以下のようになる。従来の装置では、すべてのリンク先に対してアクセスを行い、文書を取得する。このため、１文書あたりｎ個のリンク先があり、ｍ階層までアクセスを行うとすると、ｍ階層目の文書個数は、ｎのｍ乗個となり、探索の深さを大きくすると、膨大な量の文書をアクセスすることになる。

【００５２】これに対し、本発明の情報収集装置では、リンク先絞り込み条件によりリンク先を絞り込むため、アクセスする文書数を削減できる。具体的には、１文書当たり、１／ｋに絞り込まれたとすると、ｍ階層目にアクセスする文書数は、１／（ｋのｍ乗）に削減される。例えば、１／３に絞り込まれたとすると、５階層目では、絞り込みを行わない場合に比べ、１／２４３に削減

される。

【図面の簡単な説明】

【図１】本発明の原理を説明するための図である。

【図２】本発明の原理構成図である。

【図３】本発明の情報収集装置の構成図である。

【図４】本発明の情報収集装置の動作のフローチャートである。

【図５】本発明の一実施例のハイパーリンク文書の例である。

【図６】本発明の一実施例のシソーラス辞書の例である。

【図７】本発明のアクセス管理テーブルの例である。

【図８】従来の情報収集装置の構成図である。

【図９】従来装置の処理のフローチャートである。

【符号の説明】

１　探索条件設定手段，探索条件設定部

２　絞り込み条件設定手段，絞り込み条件設定部

３　アクセス候補絞り込み手段，アクセス候補絞り込み部

４　ハイパーリンク文書アクセス手段，ハイパーリンク文書アクセス部

５　探索管理手段，探索管理部

６　アクセス管理テーブル

７　取得文書蓄積部

８　処理要求受付部

９　探索結果表示部

１０　照合対象語蓄積部

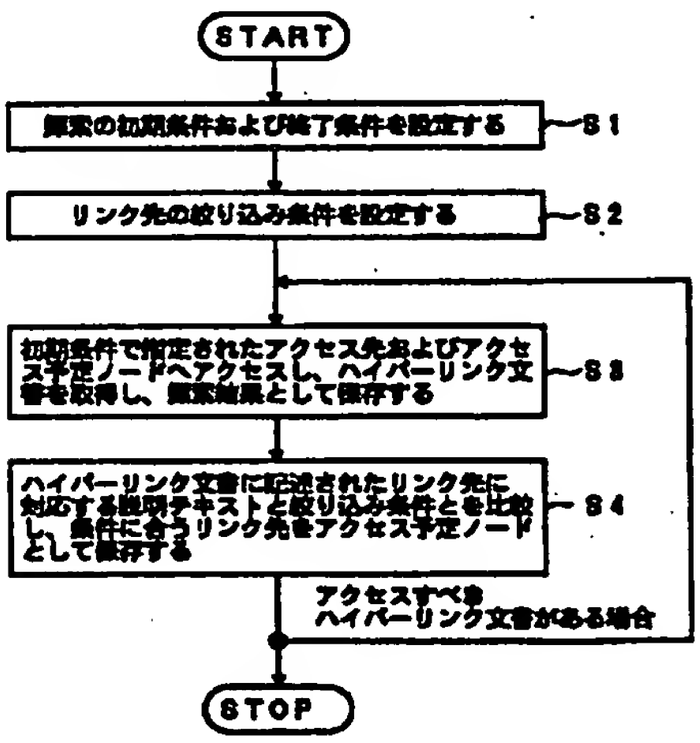
１１　シソーラス辞書

１２　ネットワーク

１３　サーバ群

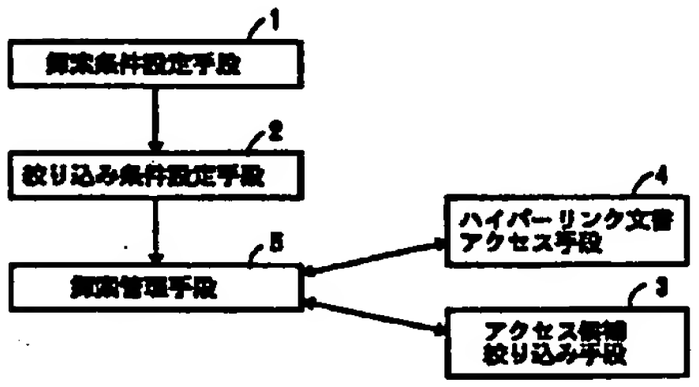
【図１】

本発明の原理を説明するための図



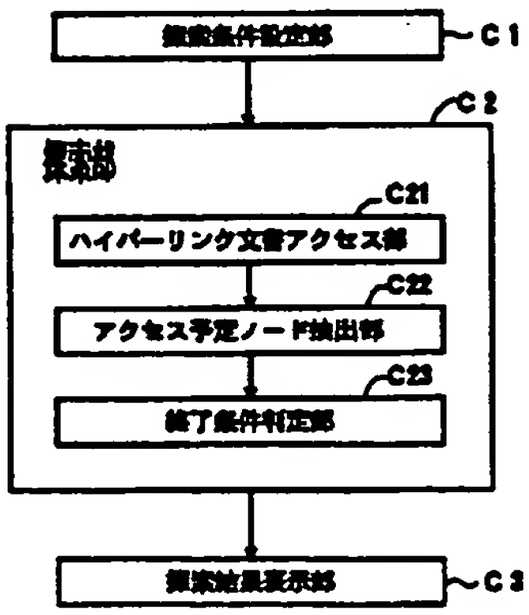
【図２】

本発明の原理構成図



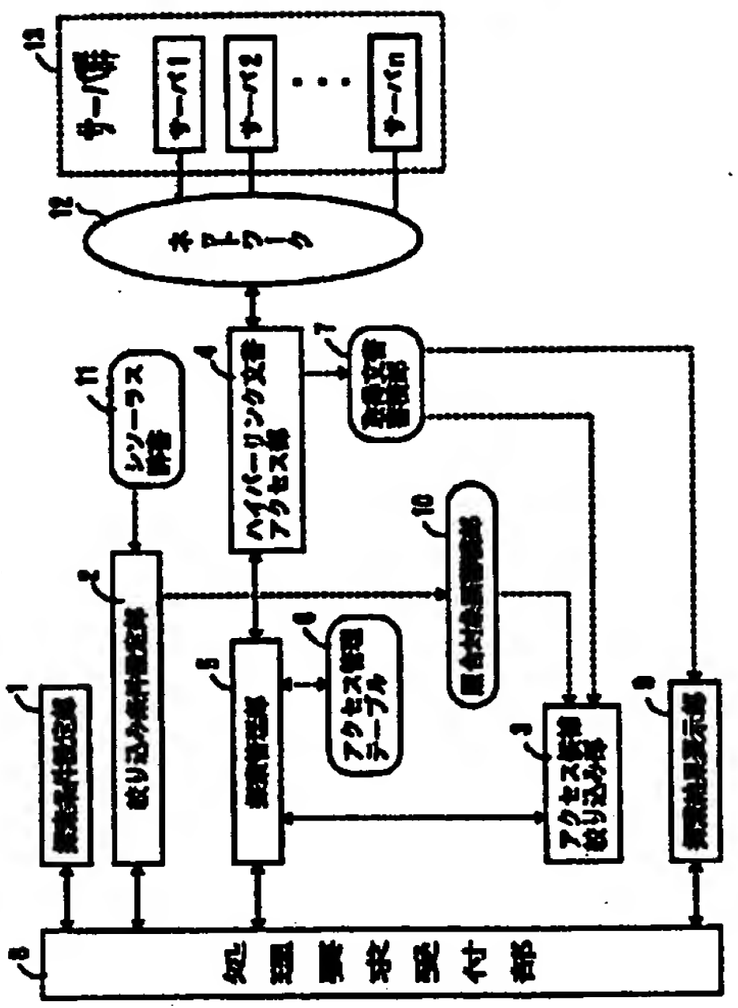
【図８】

従来の情報収集装置の構成図



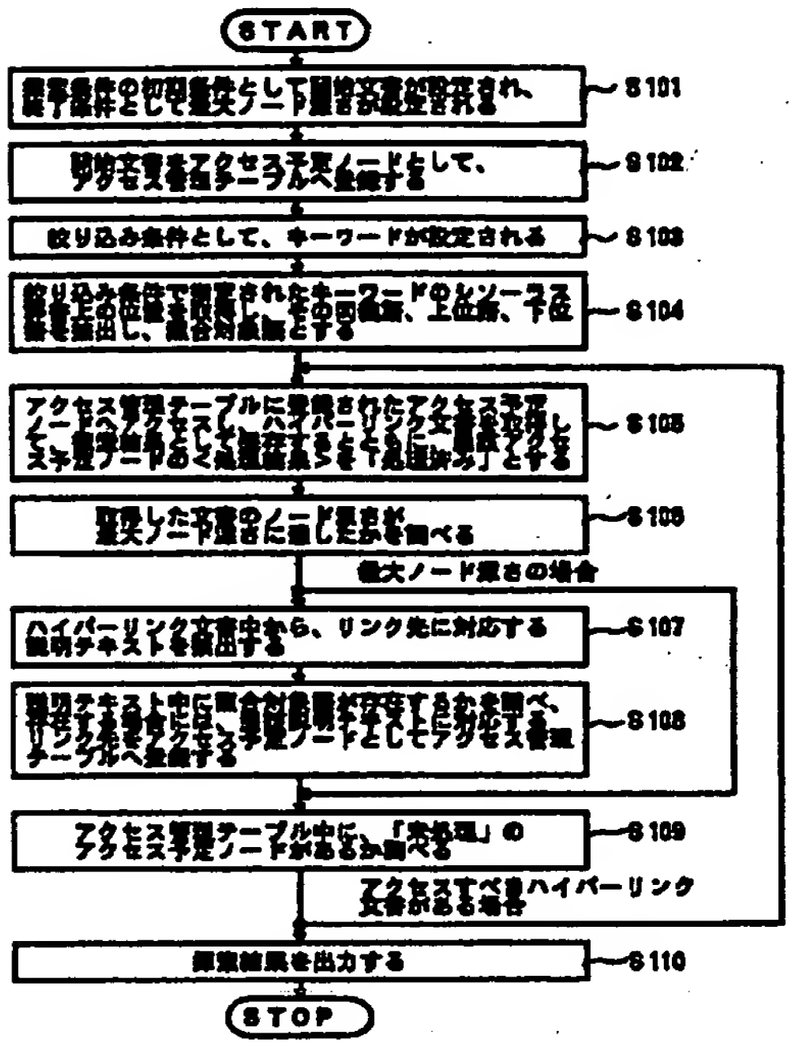
【図３】

本発明の情報収集装置の構成図



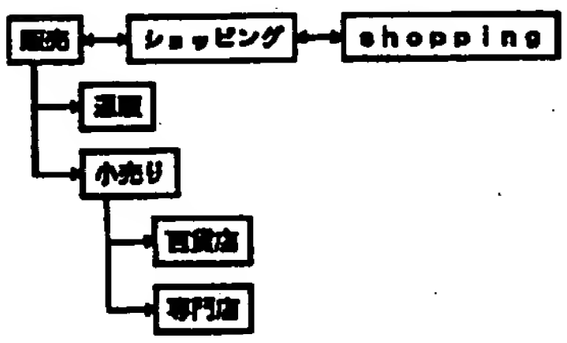
【図４】

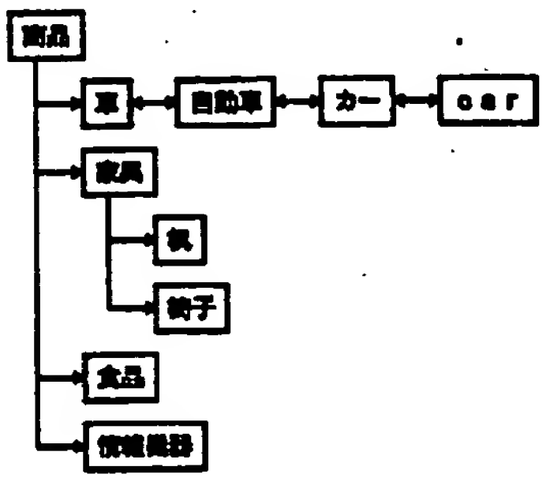
本発明の情報収集装置の動作のフローチャート



【図６】

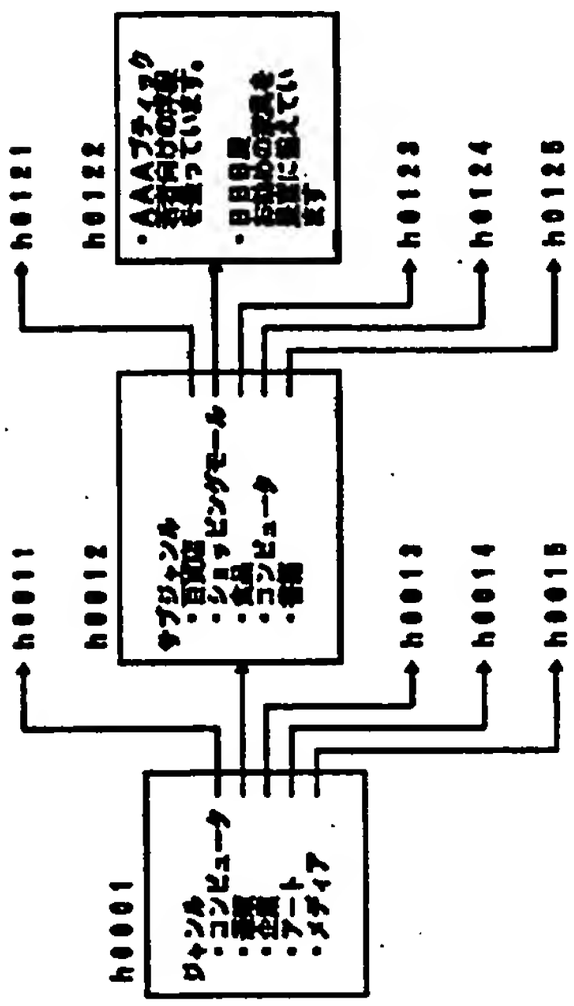
本発明の一実施例のシソーラス辞書の例





【図５】

本発明の一実施例のハイパーリンク文書の例



【図７】

本発明の一実施例のアクセス管理テーブルの例

アクセス予定ノード	処理結果	ノードの深さ
h0001	未処理	0

（a）文書アクセス前

アクセス予定ノード	処理結果	ノードの深さ
h0001	処理済み	0
h0012	未処理	1

（b）h0001文書取得後

アクセス予定ノード	処理結果	ノードの深さ
h0001	処理済み	0
h0012	処理済み	1
h0121	未処理	2
h0122	未処理	2

（c）h0012文書取得後

【図９】

検索装置の処理のフローチャート